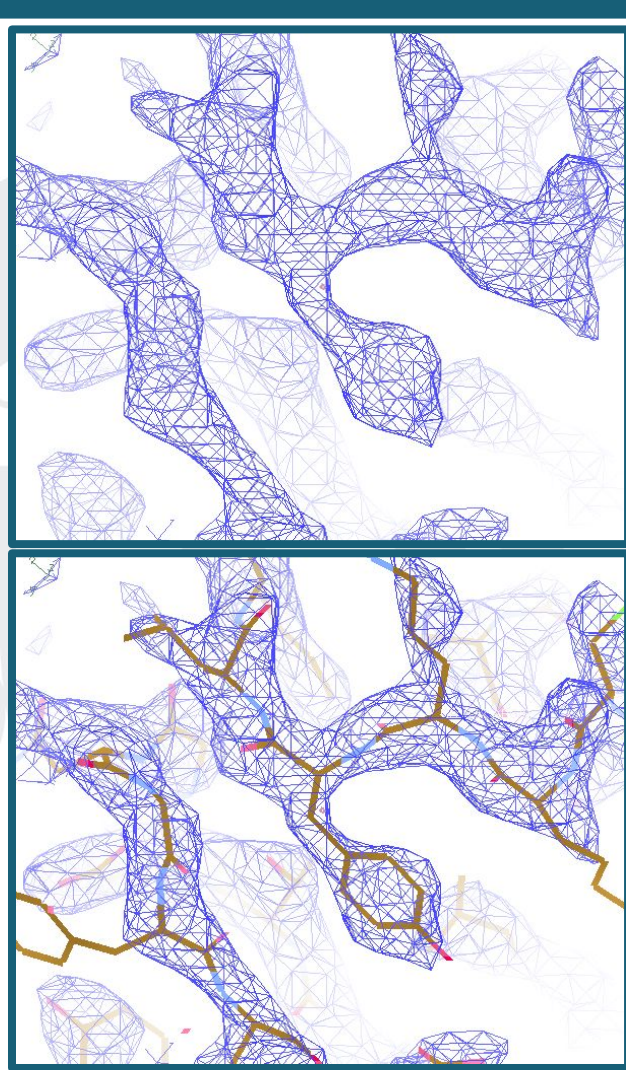


Automated model building pipeline for cryo-EM data

Soon Wen Hoh, Paul Bond, Kevin Cowtan
Icknield Model Building Workshop
9-12 October 2023



Electron cryo-microscopy



Data

1. Sample preparation
2. Imaging
3. Data acquisition, pre-processing
4. Reconstruction, map manipulation /optimisation

Model building & Refinement

Automated using ModelCraft*
- Buccaneer/Nautilus
- Servalcat[#] (REFMAC5)

Final validation & Deposition

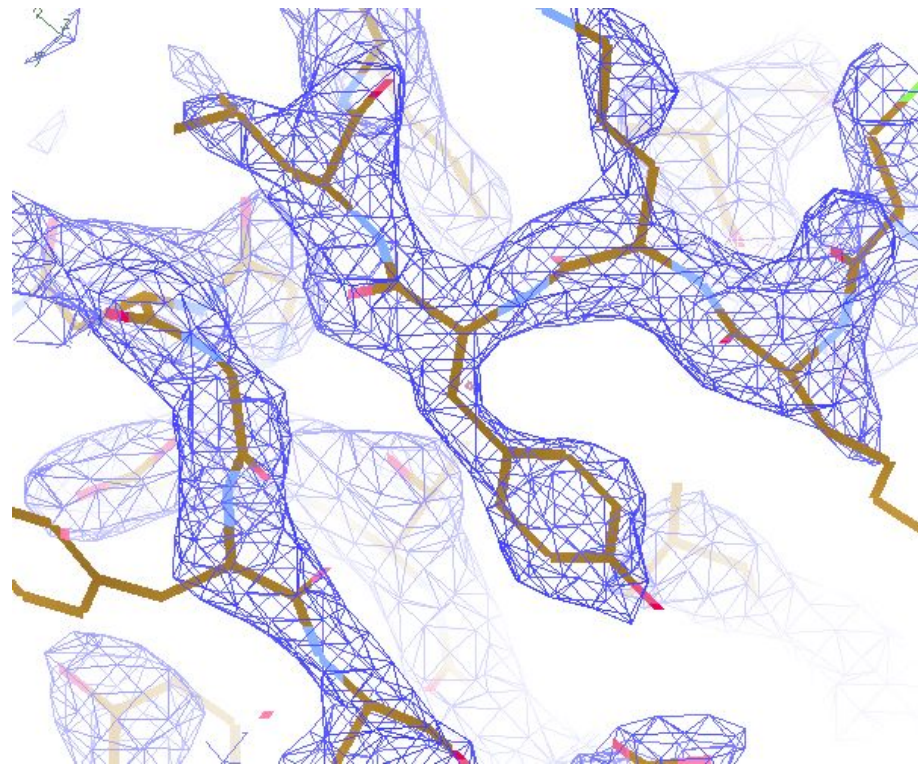
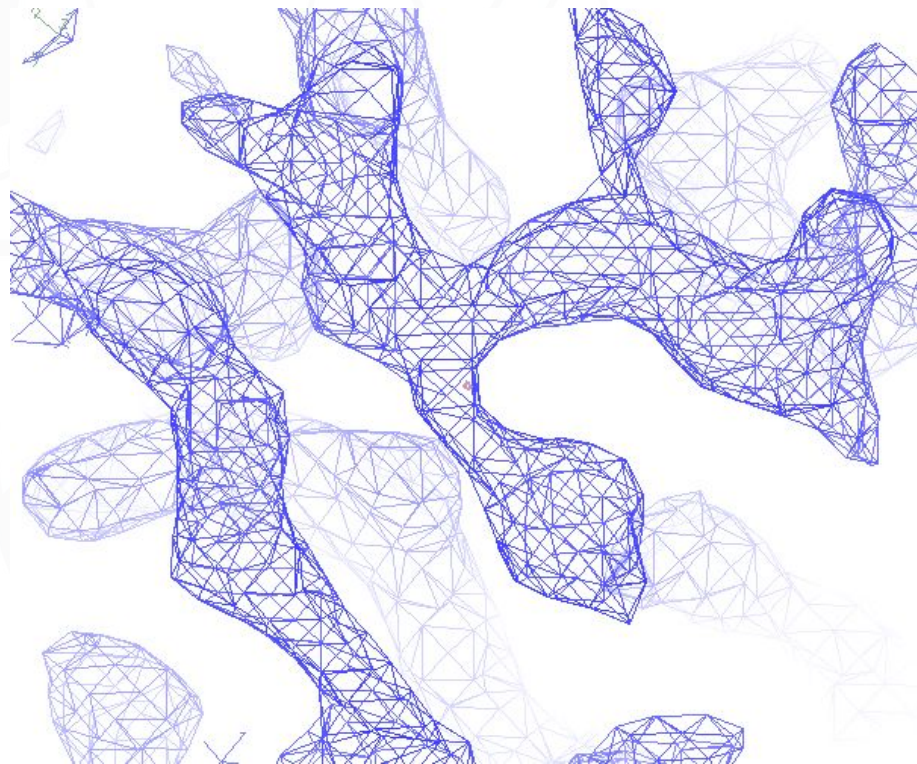
Then you find another problem to solve

Hoh et al., Acta Cryst. D76, 531-541, 2020

*Bond and Cowtan, Acta Cryst. D78, 1090-1098, 2022

[#]Yamashita, K. et al, Acta Cryst. D77, 1282-1291, 2021

Buccaneer/Nautilus - Task



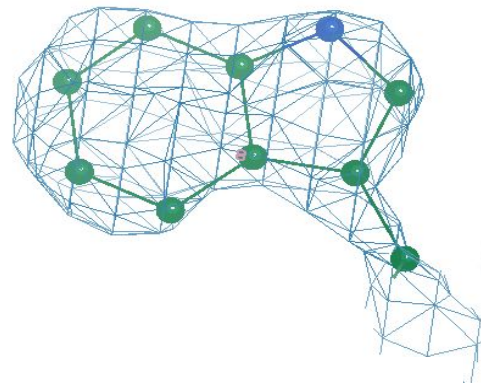
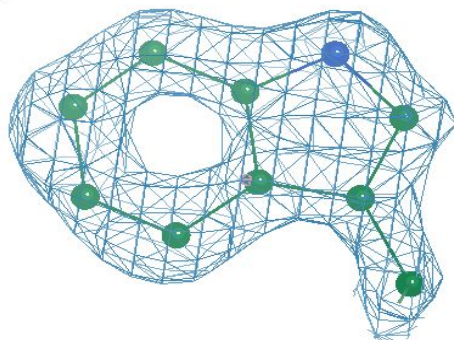
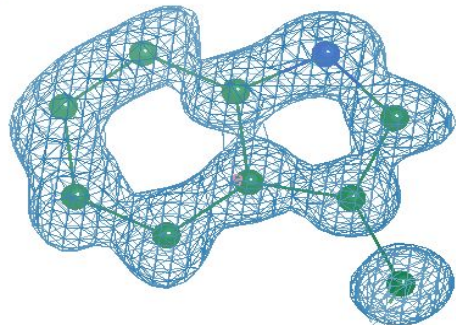
Buccaneer/Nautilus - Task

Resolution: 1Å

Resolution: 2Å

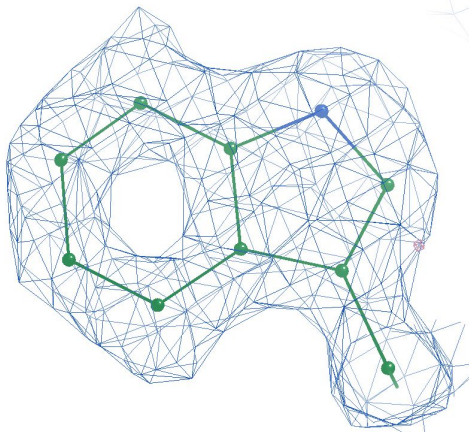
Resolution: 3Å

X-ray

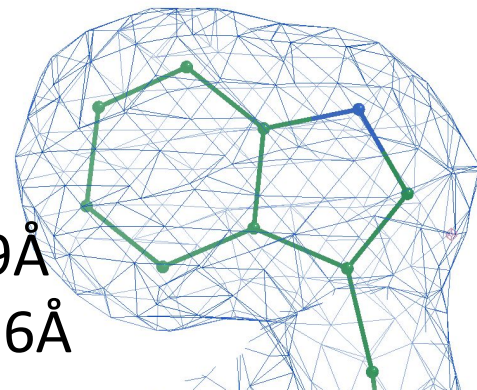


EM

Resolution
Overall : 1.75Å
Local : ~2.4Å

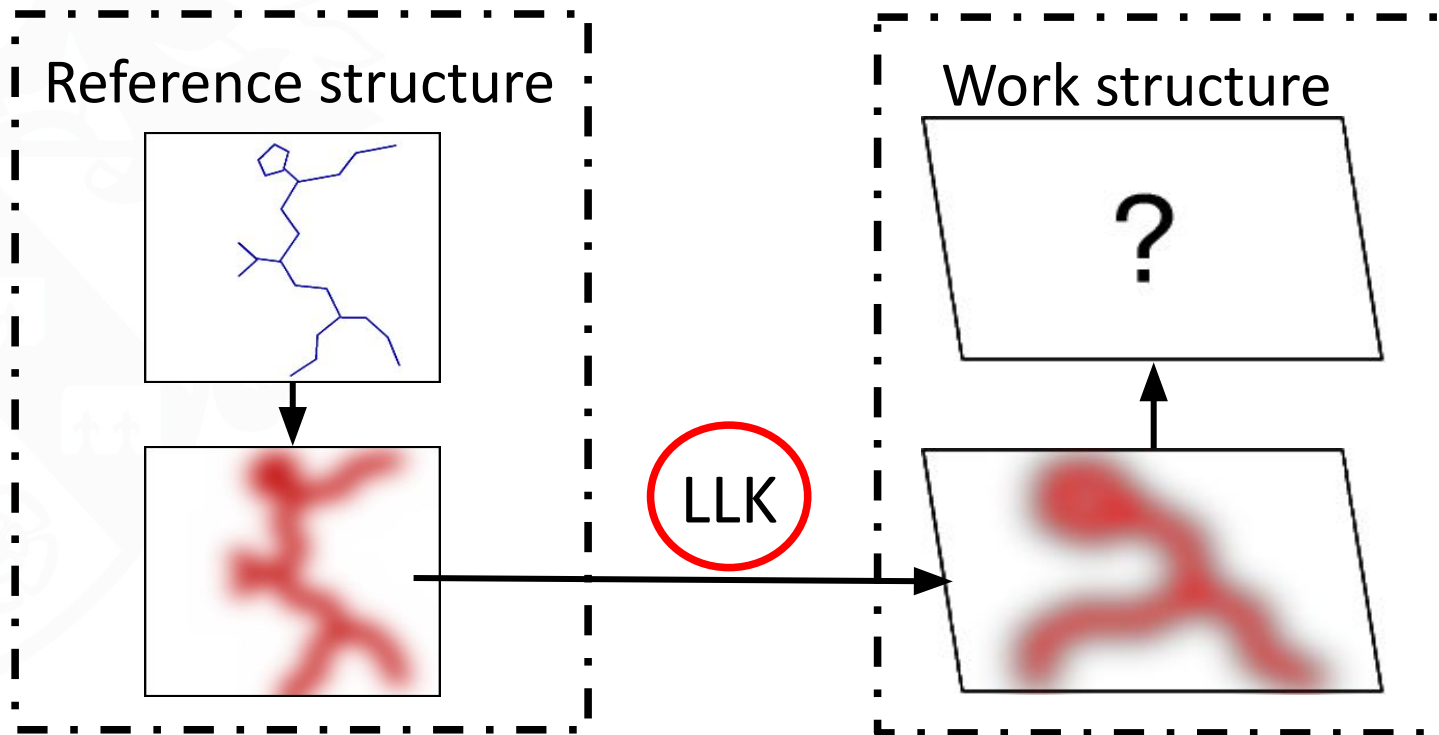


Resolution
Overall : 1.9Å
Local : ~3.6Å



Buccaneer - LLK Target function

Compare simulated map and reference model to obtain likelihood target then search for this target in the unknown map

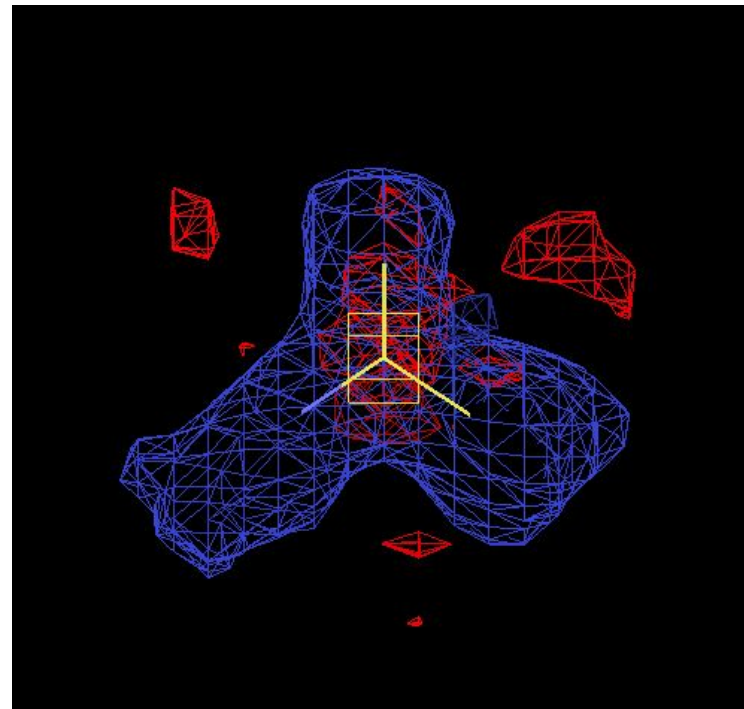
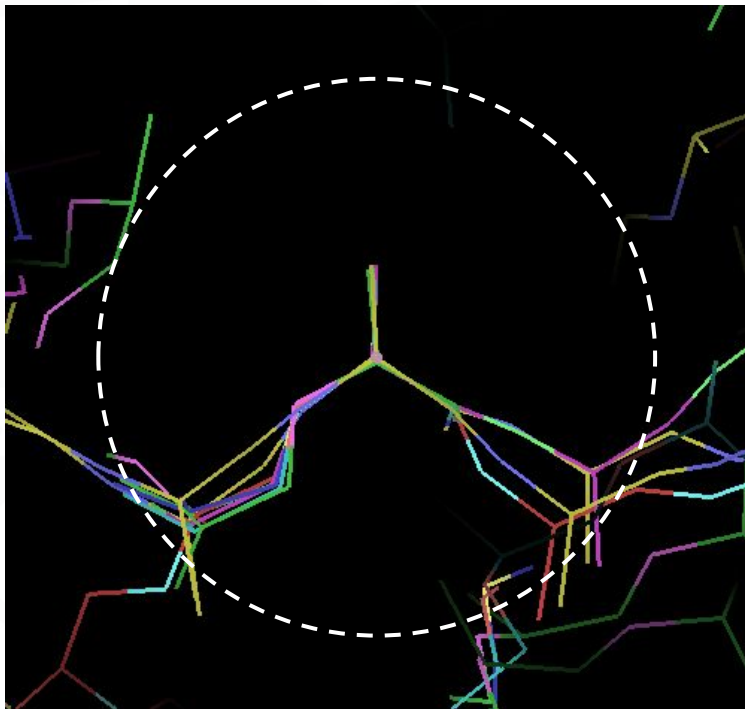


Buccaneer - LLK Target function



UNIVERSITY
of York

Measure mean and variance of reference map in a 4Å sphere around C α



Buccaneer - LLK Target Function



UNIVERSITY
of York

Likelihood functions based on conserved density features

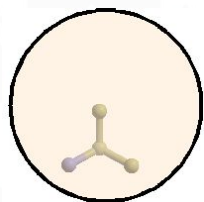
Finding, Growing

$\text{C}\alpha$ environment (4.0Å sphere)

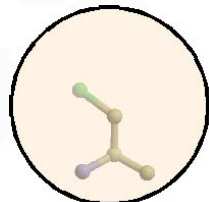


Sequencing

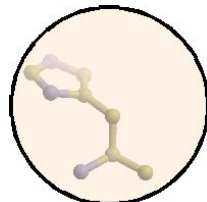
$\text{C}\beta$ environment (5.5Å sphere)



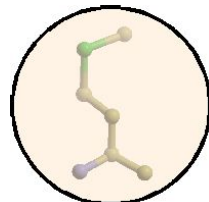
ALA



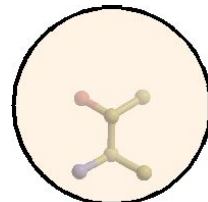
CYS



HIS



MET



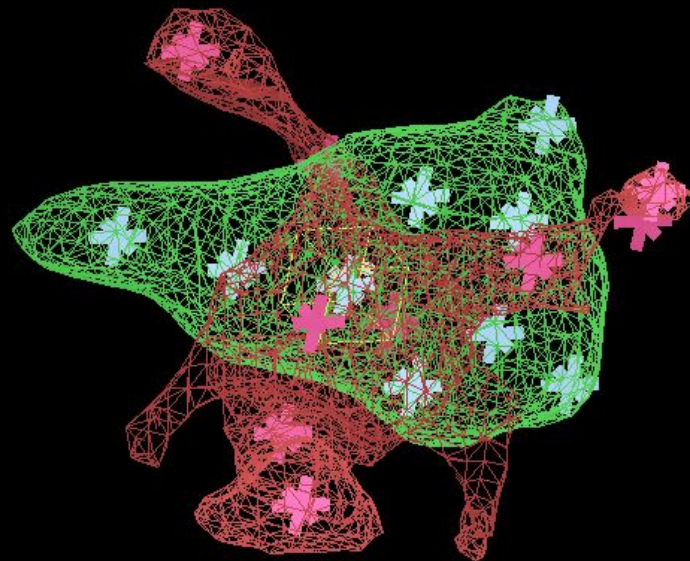
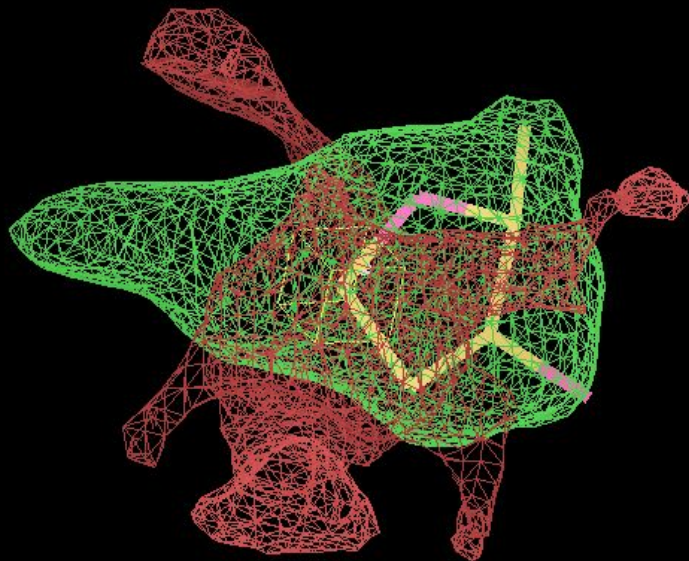
THR

... x20

Nautilus - Fingerprints



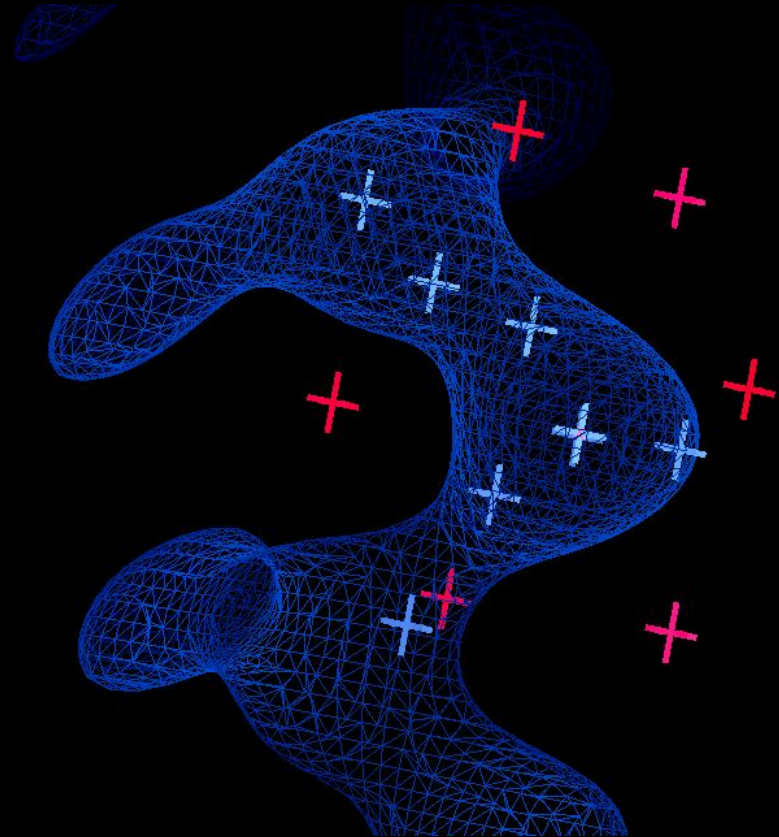
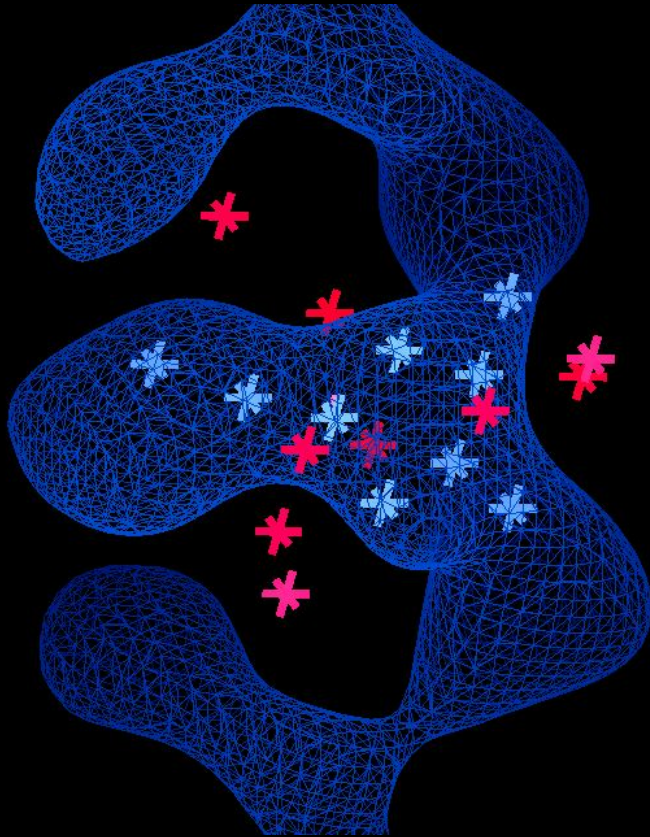
UNIVERSITY
of York



Nautilus - Fingerprints



UNIVERSITY
of Work



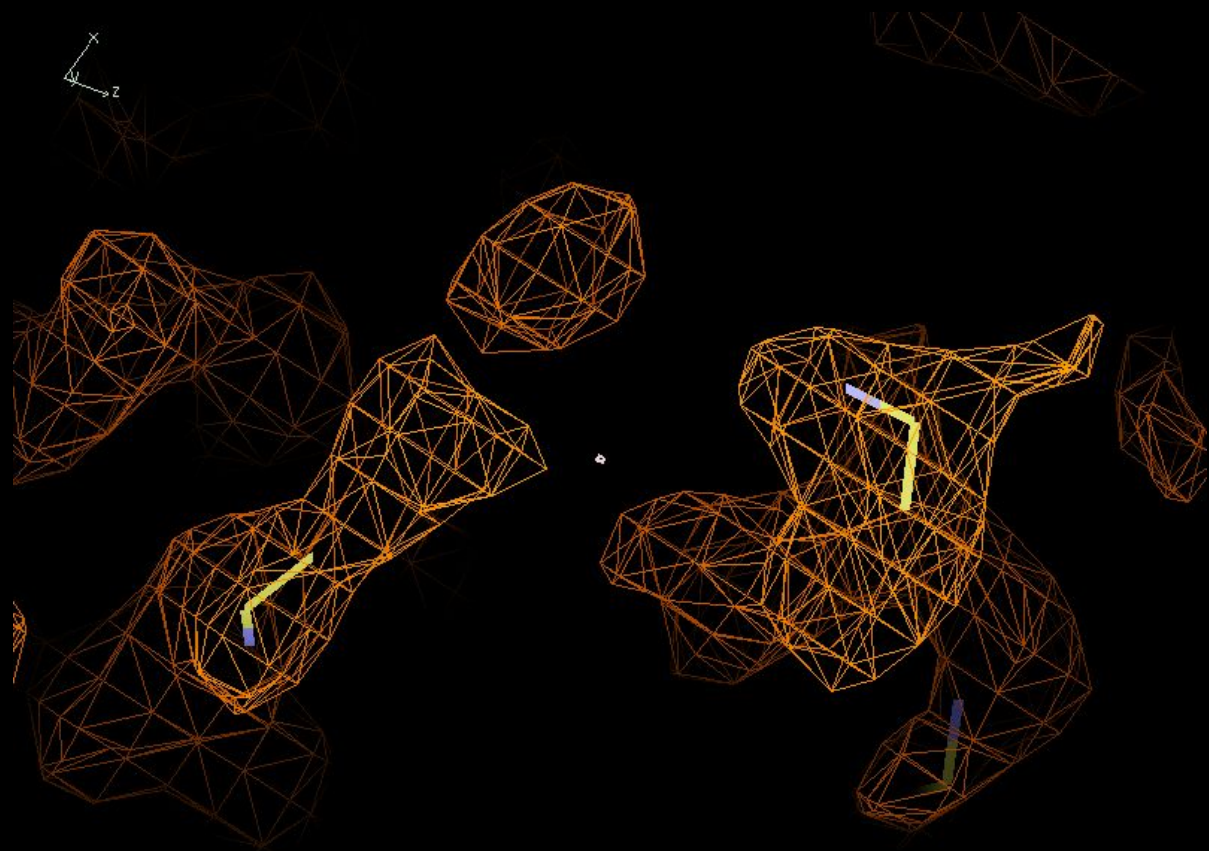
Buccaneer - Steps



1. **Find** oriented residue positions
2. **Grow** each residue into a chain fragment
3. **Join** overlapping fragments and resolve branches
4. **Link** nearby termini
5. **Sequence** the chains
6. **Correct** insertions and deletions
7. **Filter** to remove short chains
8. **NCS** superposition to extend chains
9. **Prune** residues to resolve clashes
10. **Rebuild** side chains

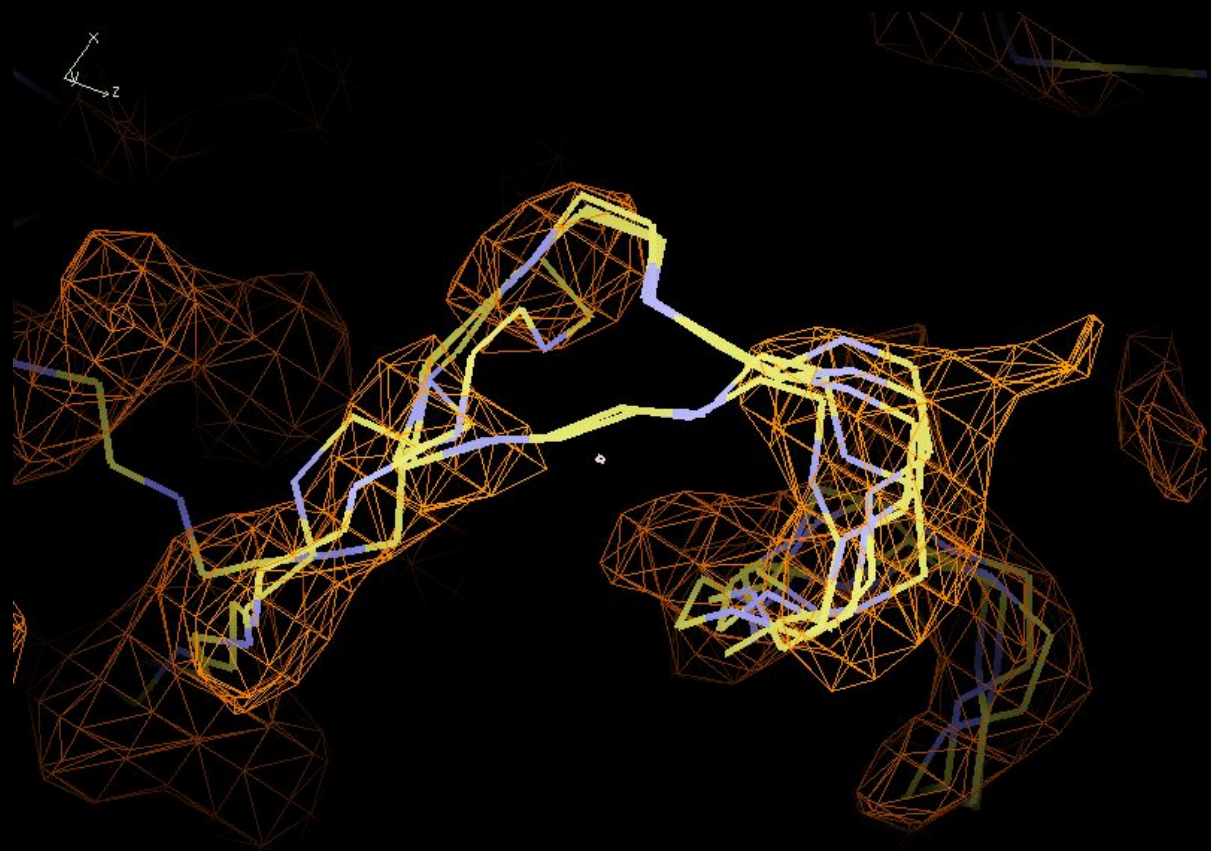
Buccaneer - Example

1. Find
2. Grow
3. Join
4. Link
5. Sequence
6. Correct
7. Filter
8. NCS
9. Prune
10. Rebuild



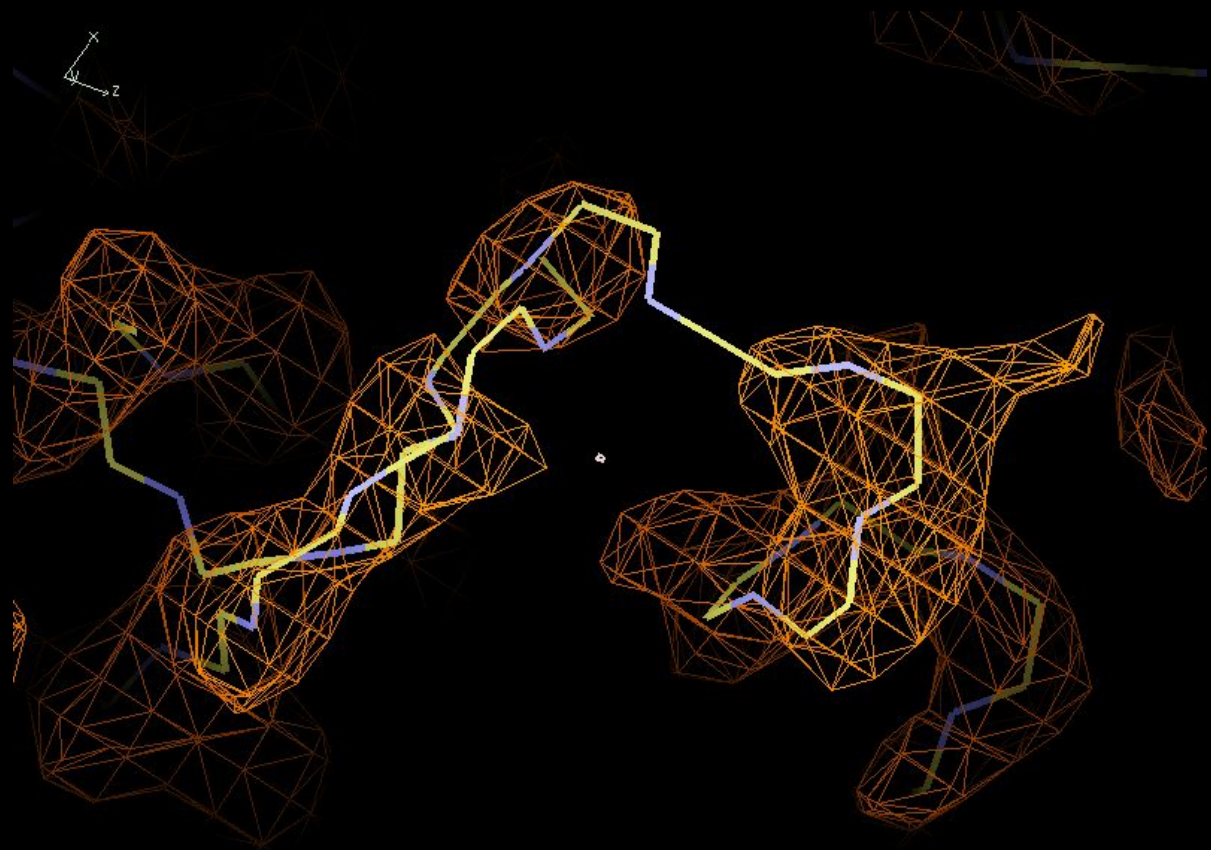
Buccaneer - Example

1. Find
- 2. Grow**
3. Join
4. Link
5. Sequence
6. Correct
7. Filter
8. NCS
9. Prune
10. Rebuild



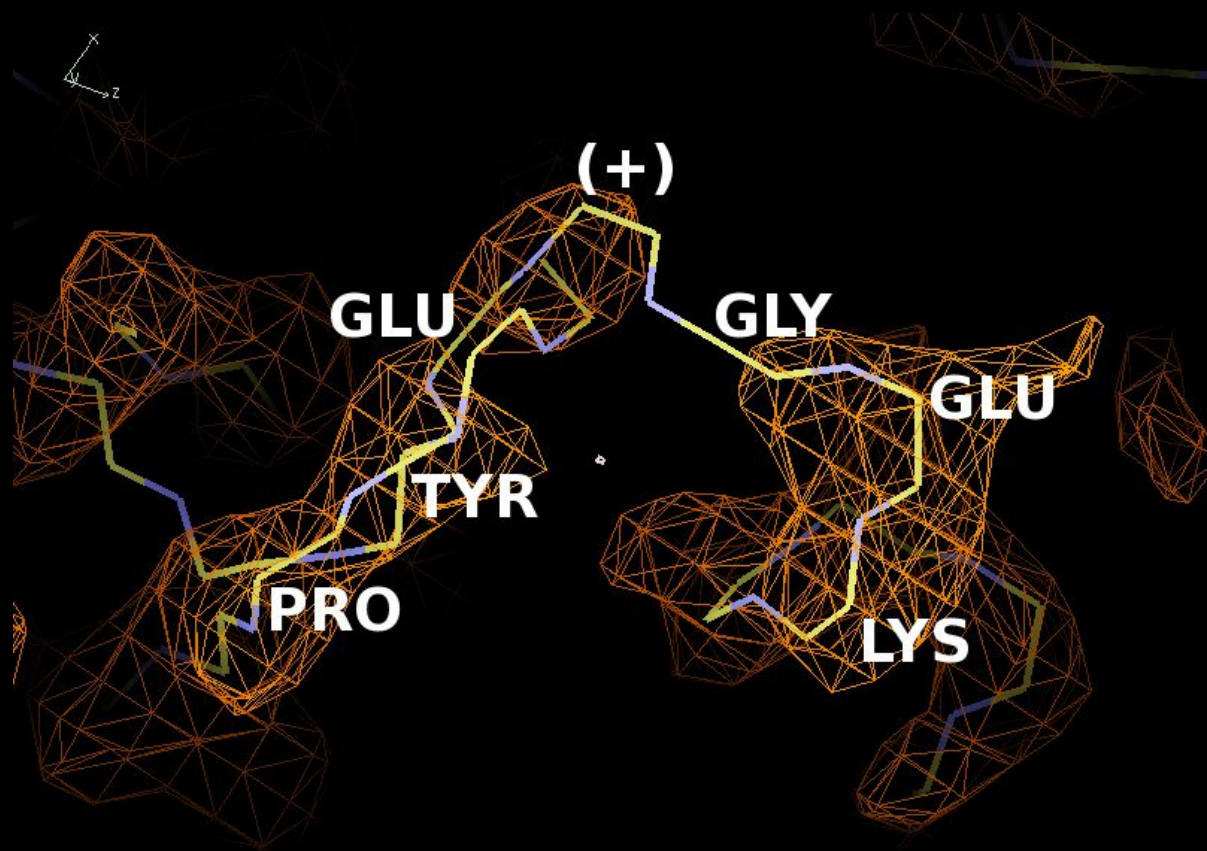
Buccaneer - Example

1. Find
2. Grow
- 3. Join**
4. Link
5. Sequence
6. Correct
7. Filter
8. NCS
9. Prune
10. Rebuild



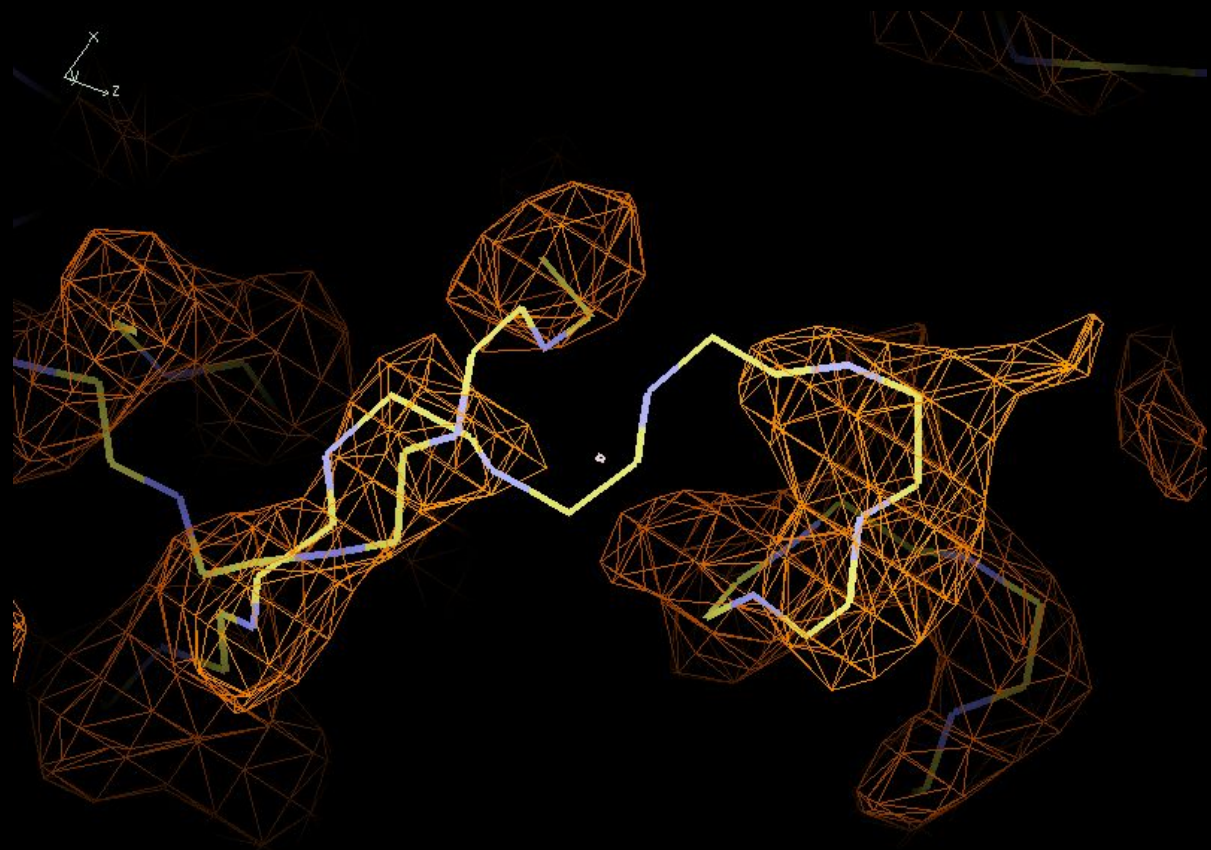
Buccaneer - Example

1. Find
2. Grow
3. Join
4. Link
- 5. Sequence**
6. Correct
7. Filter
8. NCS
9. Prune
10. Rebuild



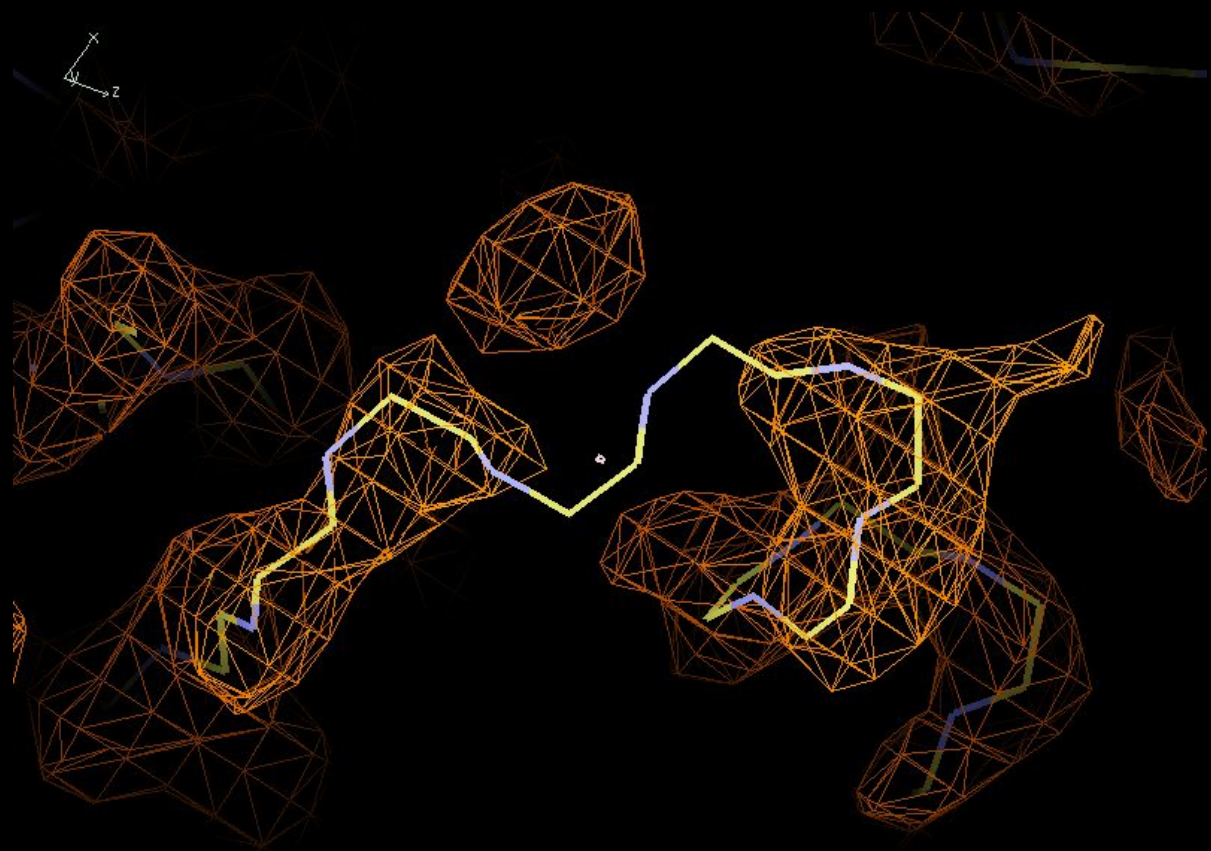
Buccaneer - Example

1. Find
2. Grow
3. Join
4. Link
5. Sequence
- 6. Correct**
7. Filter
8. NCS
9. Prune
10. Rebuild



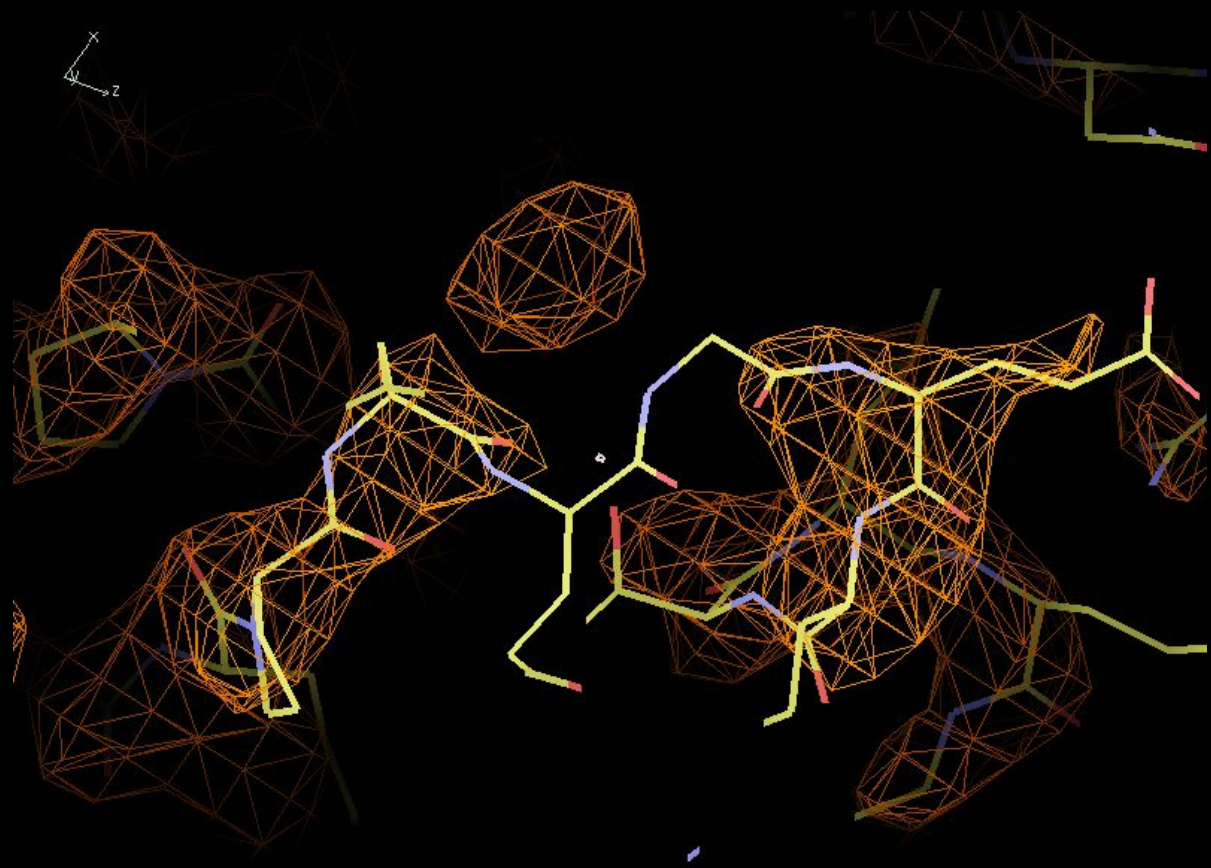
Buccaneer - Example

1. Find
2. Grow
3. Join
4. Link
5. Sequence
6. Correct
7. Filter
8. NCS
- 9. Prune**
10. Rebuild



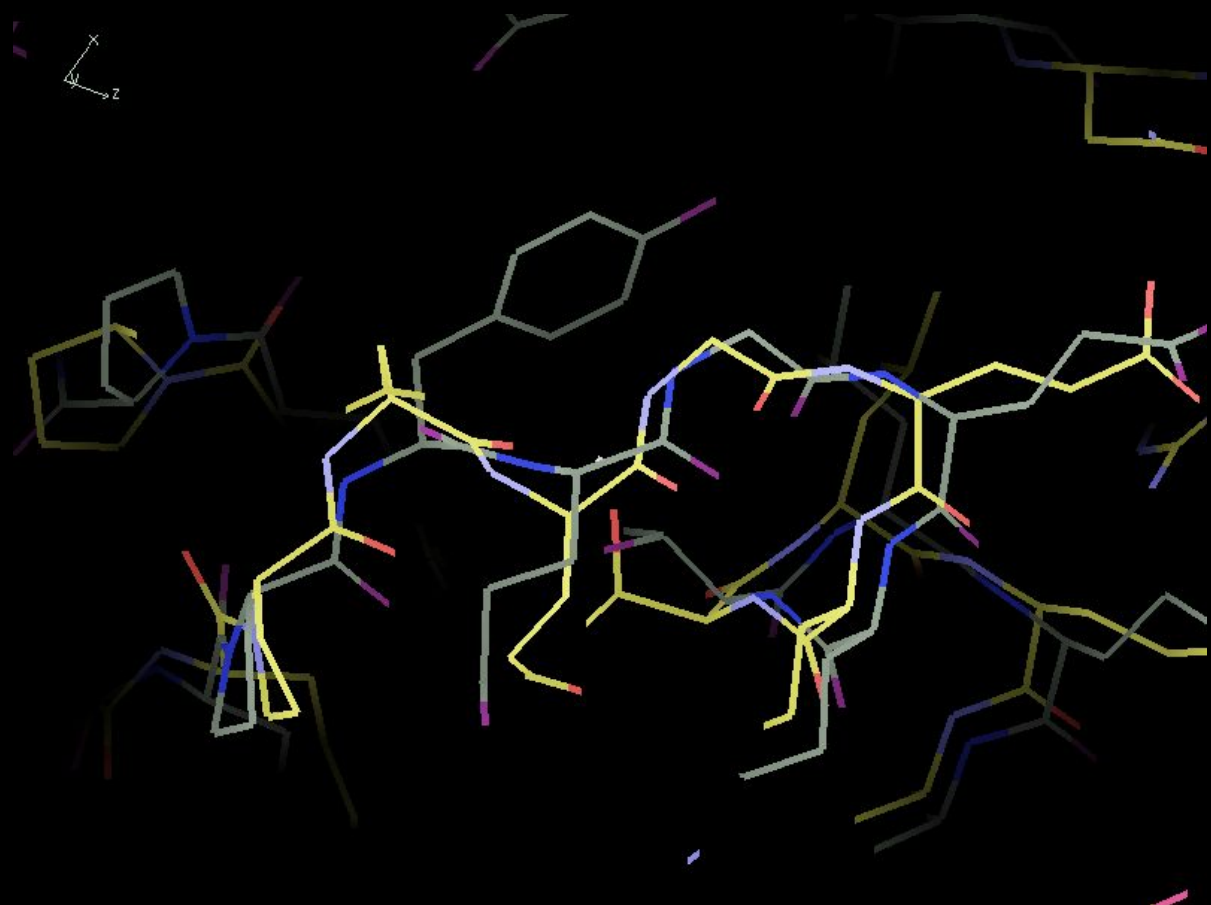
Buccaneer - Example

1. Find
2. Grow
3. Join
4. Link
5. Sequence
6. Correct
7. Filter
8. NCS
9. Prune
- 10. Rebuild**



Buccaneer - Example

1. Find
2. Grow
3. Join
4. Link
5. Sequence
6. Correct
7. Filter
8. NCS
9. Prune
10. Rebuild



Buccaneer - CCP-EM Pipeline v1.#



Buccaneer

1. **Find** oriented residue positions
2. **Grow** each residue into a chain fragment
3. **Join** overlapping fragments and resolve branches
5. **Sequence** the chains
10. **Rebuild** side chains

5x

Servalcat/Refmac

Refine coordinates and B-factors

20x

Repeat n cycles

ModelCraft - Pipeline



UNIVERSITY
of York

Shift-field refinement (*Sheetbend*) - *Refmac*

Xtal

1. Prune residues (*Coot*) - *Refmac*

Xtal

2. Density modification (*Parrot*)

Xtal

3. Add dummy atoms (*Coot*) - *Refmac*

Xtal

4. Build protein (*Buccaneer*) - *Refmac/Servalcat*

Xtal/EM

5. Prune chains (*Coot*) - *Refmac*

Xtal

6. Build RNA & DNA (*Nautilus*) - *Refmac/Servalcat*

Xtal/EM

7. Add waters (*Coot*) - *Refmac*

Xtal

≤ 25x

Rebuild side chains (*Coot*) - *Refmac*

Xtal

CCP-EM Doppio

Required: either
a full map or
two half maps

Required













Optional

~~ModelCraft~~ ModelCraft Task plugin v1

RUN JOB INFO RESET PARAMETERS

Job alias:

Main

Job title	<input type="text"/>	
Input map 1 (half map 1 or full map) *	required*	
Input map 2 (half map 2)	<input type="text"/>	
Input map resolution *	-1	
Input sequence *	required*	
Input map mask	<input type="text"/>	
Starting model	<input type="text"/>	
Max cycles of build-refine runs	25	
Auto stop cycles without improvements	4	
B-factor to sharpen or blur	0	
Ligand restraint dictionary file	<input type="text"/>	 

CCP-EM Doppio

RESULTS

LOGS

I/O

PARAMS

Job 30 - Buccaneer

Residues built and FSC average for modelcraft.cif

Residues built	FSC average
573	0.8177

Overlaid map: Import/job023/3488_run_half1_class001_unfil.mrc model: Buccaneer/job030/modelcraft.cif

Volume

Import/job023/3488_run_half1_class001_u...

Isosurface 2.00 σ

Opacity 0.5

Iso Value 2

Atomic Models

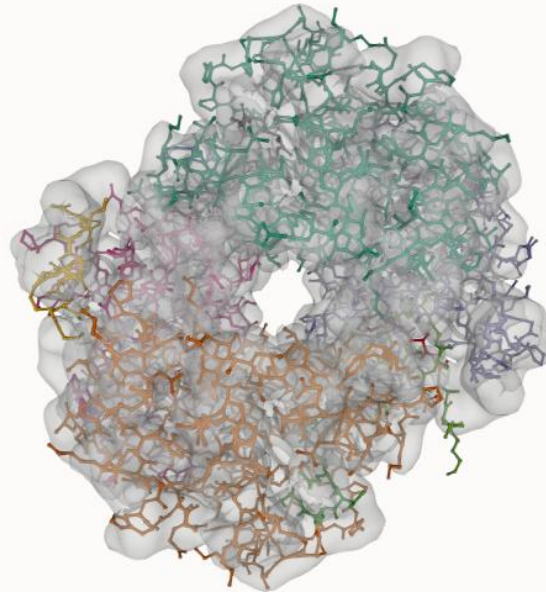
All Ball & Stick

Type Ball & Stick

Color Theme Chain Id

Size Theme Physical

Nothing to Update



ModelCraft - Command Line



UNIVERSITY
of York

```
$ modelcraft --version  
3.4.0
```

```
$ modelcraft em \  
> --contents input_seq.fasta \  
> --map full_map.mrc \  
> --model starting_model.cif \  
> --mask input_mask.mrc
```

```
$ modelcraft em --help
```

```
$ modelcraft --version  
3.4.0
```

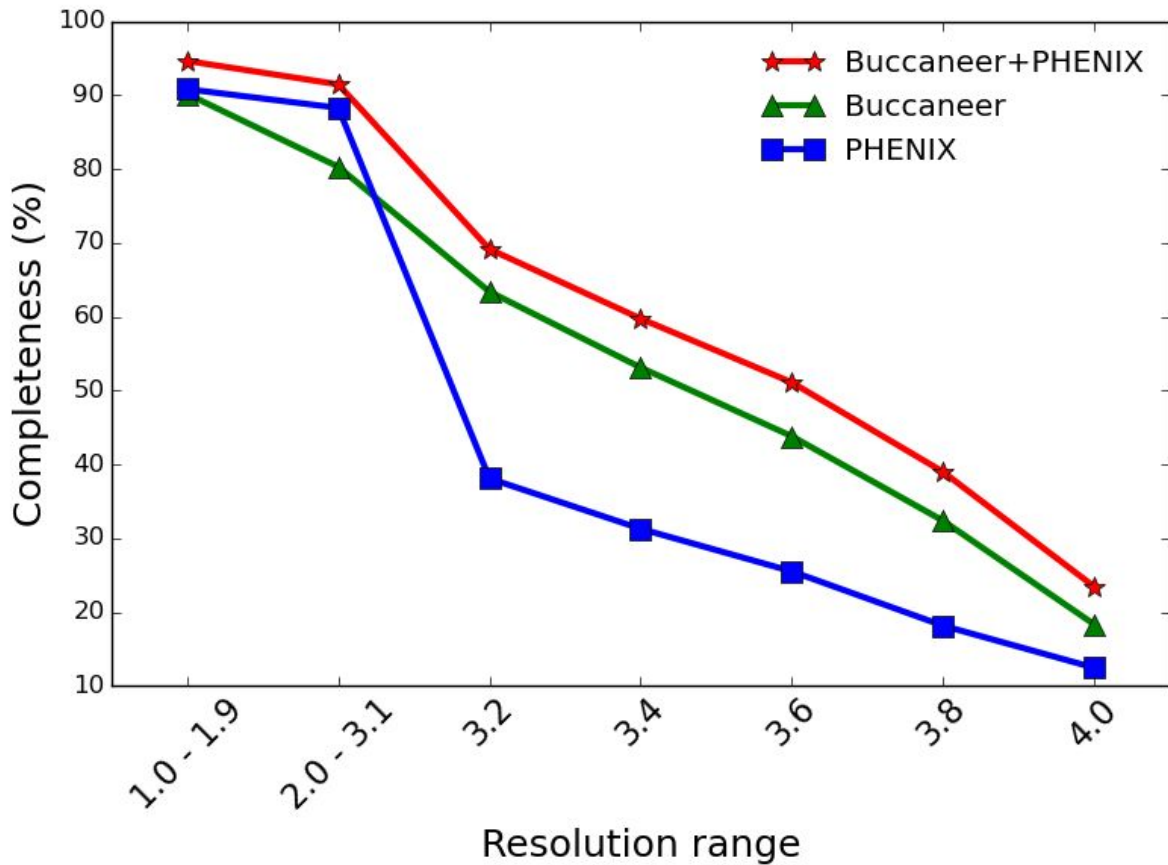
```
$ modelcraft em \  
> --contents input_seq.fasta \  
> --map half_map1.mrc half_map2.mrc \  
> --model starting_model.cif \  
> --mask input_mask.mrc
```

```
$ modelcraft em --help
```

Combination of Model Building Programs



Alharbi, E. et al.,
Acta Cryst. D79,
326-338, 2023



Acknowledgements

Kevin Cowtan

Eleanor Dodson

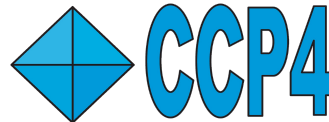
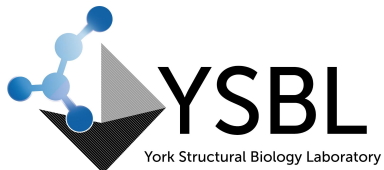
Paul Bond

Keitaro Yamashita (Servalcat)

Agnel P. Joseph Colin Palmer

Matt Iadanza Tom Burnley

Marcin Wojdyr (GEMMI)



Biotechnology and
Biological Sciences
Research Council